**QUIZZZZZZ............**

You have 60 minutes to complete the following 3 questions. You may use your notes and slides but nothing to connect to the internet. Good luck.

1. Assume the following fragments comprise your document collection :

Doc 1 : banking on banks to raise the interest rate

Doc 2 : jogging along the river bank to look at the sailboats

Doc 3 : jogging to the bank to look at the interest rate

Doc 4 : buzzer-beating shot banked in !

Doc 5 : scenic outlooks on the banks of the Potomac River

Construct the vector space term-document matrix for the above documents using ***tf.idf*** term weighting. Normalize your vectors. Assume you drop stopwords and do the stemming. Simulate the retrival documents in response to the following queries (a) bank interest and (b) bank river. Indicate the order in which documents will be retrieved including the associated similarity scores (40%).

1. Two retrieval engines A and B index the same document collection. Each engine return the top-30 documents for a single query. The following listings denote relevant documents with a “+” and no-relevant ones with a “-“ :

A : + + - + - - - + - - - + - - - - - - - + - - - - - - - - - + ; (7 “+” dan 23 “-“)

B : - + + - - - + - - + + - + - - - + - - + - - + - + - - - - - ; (10 “+” dan 20 “-“)

The collection contains 12 relevant documents for the query.

1. Draw the plot for precision-recall for both engines (20%)
2. Which engines returns better results ? Explain. (10%)
3. Using SVD and cosine similariy, which of the following images (I1 and I2) is very much similar to Itest. Matlab may be used to help you to calculate the SVD of the document matrix A (30%).

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | 0 | 0 |  | 0 | 5 | 0 |  | 0 | 7 | 3 |
| 5 | 5 | 5 |  | 0 | 5 | 0 |  | 0 | 5 | 5 |
| 0 | 0 | 0 |  | 0 | 5 | 0 |  | 0 | 5 | 0 |
| IMAGE I1 | | |  | IMAGE I2 | | |  | IMAGE Itest | | |